



LATHAM &amp; WATKINS LLP

MORRISON FOERSTER

January 13, 2025

VIA ECF

Hon. Ona T. Wang  
 United States District Court  
 500 Pearl Street  
 New York, NY 10007

Re: **Discovery dispute concerning custodians and date ranges**

Dear Judge Wang:

OpenAI requests a pre-motion conference to address The New York Times's deficient designation of custodians and refusal to collect pre-2022 documents for key custodians. The Times is a publicly-traded company with a market capitalization in the billions, thousands of employees, and a technology and engineering team which Chief Technology Officer Jason Sobel has described as "second only to the newsroom" in size.<sup>1</sup> The Times is also a sophisticated litigant, and has been party to hundreds of lawsuits in the past ten years. Yet, despite The Times's significant resources and litigation experience, it has shirked its discovery responsibilities by failing to add custodians likely to have responsive information, failing to timely provide hit counts, and failing to apply appropriate date ranges to its collection.

The Times's refusal to comply with its discovery obligations is prejudicing OpenAI's ability to advance its defense, including preparing for depositions, which The Times claims it is eager to commence. OpenAI therefore: (1) reasserts its request for The Times to add David Rubin, Michael Greenspon, Hannah Masuga, and Sam Dolnick as custodians (ECF 316); (2) further requests the addition of Justin Heideman, Staff Software Engineer, and Elijah Soria, Senior Software Engineer as custodians; and (3) requests that The Times collect and search custodial files back to 2018 for RFPs related to OpenAI's alleged misconduct and The Times's allegations of harm.

**I. The Times should add Rubin, Greenspon, Masuga, and Dolnick as custodians.**

The Times does not dispute that Rubin, Greenspon, Masuga and Dolnick are likely to have responsive documents. *See* ECF 330 at 1-3.<sup>2</sup> At the December 3 hearing, The Times's principal argument against adding these custodians was burden: The Times represented that, as a "small company" without an "expansive IT department," even the production of hit counts for these four individuals was beyond its capabilities and would require "paus[ing] the process of

<sup>1</sup><https://techcrunch.com/2023/02/08/how-the-nyt-is-building-a-modern-tech-stack-to-drive-every-part-of-its-media-biz/>.

<sup>2</sup> Nor could it—OpenAI's ongoing review of The Times's productions has revealed additional evidence that these requested custodians are likely to have unique, relevant information. *See e.g.*, Ex. A (NYT 00042062) [REDACTED]

[REDACTED] Ex. B (NYT 00069189); Ex. C (NYT 00067235); Ex. D (NYT 00092341) [REDACTED]

collecting documents from [agreed-upon custodians].” Ex. E (Dec. 3, 2024 Hr’g Tr.) at 39:10-11; *id.* at 40:22-24. In reliance on these representations, the Court ordered the parties to meet and confer after OpenAI had received The Times’s production from the ten custodians The Times agreed to add in October and November, and to re-raise the issue of the 4 additional custodians with the Court if not resolved. *Id.* at 42:10-25.

Notwithstanding this guidance, The Times failed to produce a single document between November 22 and January 10—one business day before motions were due in advance of the January 22 hearing.<sup>3</sup> Indeed, The Times only provided partial search term hit counts for the ten custodians for the first time on December 23, 2024, Ex. F (12/23/24 Peaslee email), and it has indicated that collection remains ongoing for these individuals. Ex. G (1/6/25 Peaslee email). In other words, The Times has yet to make meaningful custodial productions from approximately *half* of its agreed custodians and has not provided a clear timeline for when it plans to do so.

The Times cannot blame these delays on its purportedly meager technological capabilities or resources. *See Zhulinska v. Niyazov Law Grp., P.C.*, Case No. 21-CV-1348, 2021 WL 5281115, at \*3 (E.D.N.Y. Nov. 12, 2021) (a party cannot evade its discovery obligations by offering “merely speculative” concerns regarding burden). Although The Times told the Court that it does not operate like “a software company,” Ex. E at 40:23, CTO Jason Sobel has boasted that “The Times does engineering much like any company folks would call ‘a tech company’” and has “a cross-functional team of designers, engineers and product managers . . . to ship websites and apps and back-end tech.”<sup>4</sup> Moreover, according to its SEC filings, the Times has “[i]n the third quarter and first nine months of 2024, . . . recorded \$4.6 million and \$7.6 million, respectively, of pre-tax litigation-related costs in connection with” this lawsuit.<sup>5</sup> This represents a small fraction of the Times’s overall revenues.<sup>6</sup> If The Times has limited its custodial collections and searches, it is not because it lacks the means to conduct these routine discovery tasks.

Given the Times’s production delays, OpenAI cannot wait months for The Times to provide hit counts for the four proposed custodians before re-raising this issue. The Court should order The Times to add Rubin, Greenspon, Dolnick, and Masuga as custodians now.

<sup>3</sup> The Times’s January 10 production totaled only 355 documents, the majority of which are transcripts of earnings calls, press releases, and other financial and audience reporting, not custodial documents (The Times is listed as the custodian for 292 of these documents).

<sup>4</sup> <https://techcrunch.com/2023/02/08/how-the-nyt-is-building-a-modern-tech-stack-to-drive-every-part-of-its-media-biz/>. *See also* <https://www.theverge.com/23651323/new-york-times-ceo-meredith-kopit-levien-wordle-the-athletic-sports-cooking> (Times CEO Meredith Kopit Levien describing the Times as a “tech-enabled journalism company” with “extraordinary digital product development, engineering, and data talent.”).

<sup>5</sup> <https://d18rn0p25nwr6d.cloudfront.net/CIK-0000071691/fa55ec3e-dfdf-4da8-985a-b96f8a7da421.pdf>.

<sup>6</sup> According to the Times’s Q3 2024 10-Q, “[t]otal revenues increased 7.0% to \$640.2 million in the third quarter of 2024 from \$598.3 million in the third quarter of 2023” and “[o]perating profit increased 20.7% to \$76.7 million in the third quarter of 2024 from \$63.6 million in the third quarter of 2023.” *Id.*

**II. The Times should add Justin Heideman and Elijah Soria as custodians.**

OpenAI further seeks the addition of Justin Heideman, Staff Software Engineer, and Elijah Soria, Senior Software Engineer, as custodians for search terms aimed at the operation of The Times’s paywall, webmaster controls, and exclusion protocols intended to prevent web crawlers from accessing its works. The Times concedes that such discovery is relevant, *see* ECF 330 at 3, Ex. H (Times Rsp. to RFPs 38, 133, 166), and has specifically identified Heideman and Soria as knowledgeable regarding these subjects. Ex. I (Times Rsp. to MSFT Rgs 7 and 8).

Yet, The Times’s production to date regarding these issues has been paltry. *See* Ex. G (1/2/25 Orman email) (noting that The Times has not yet produced “documents sufficient to show the time periods during which The Times implemented Exclusion Protocols on [its] website.”). The issue can be traced to The Times’s designation of custodians; while The Times has agreed to search the files of high-level custodians<sup>7</sup> for such documents, it refuses to search the files of Heideman and Soria, engineers who are more likely to have day-to-day discussions about how The Times’s access controls operate. *See* Ex. I; Ex. J (1/3/25 Peaslee email). Heideman and Soria should be added as custodians for these specific issues.

**III. The Times should expand the date range for search terms aimed at communications regarding OpenAI’s alleged misconduct and The Times’s allegations of harm.**

Finally, OpenAI requests that The Times be ordered to collect custodial documents back to 2018 for search terms aimed at communications regarding OpenAI’s alleged misconduct and The Times’s claims of harm. The Times originally applied a date range limitation of November 2022 to May 2024 to its entire custodial collection, even though it bases its claims on alleged misconduct pre-dating November 2022 and has sought extensive discovery from OpenAI regarding LLMs released from 2018 onwards. The Times has since conceded that January 2018 to May 2024 is the appropriate timeframe for search terms aimed at OpenAI’s alleged misconduct, but has only agreed to apply that date range to a handful of custodians’ files.

This limitation is improper; OpenAI is entitled to communications from all relevant custodians reflecting their awareness of, and views regarding, OpenAI’s alleged misconduct prior to November 2022. As for harm-related search terms, The Times cannot at the same time maintain damages claims based on OpenAI’s pre-2022 conduct, yet prevent OpenAI from obtaining the documents that would allow it to challenge these assertions. Notably, The Times is the only News plaintiff which has imposed such limitations—the Daily News plaintiffs and the Center for Investigative Reporting have both agreed to conduct custodial collections back to 2018. The Times should be ordered to do so as well.

---

<sup>7</sup> The Times has agreed to search the files of custodians such as its Chief Data Officer, CTO, Head of Platform Strategy, and, for a limited timeframe, Senior Director of Technical Search for such documents.

KEKER  
VAN NEST  
& PETERS

LATHAM & WATKINS LLP

MORRISON & FOERSTER

Sincerely,

KEKER, VAN NEST &  
PETERS LLP

*/s/ Michelle S. Ybarra*

---

Michelle S. Ybarra

LATHAM & WATKINS LLP

*/s/ Elana Nightingale  
Dawson*

---

Elana Nightingale Dawson

MORRISON & FOERSTER  
LLP

*/s/ Rose S. Lee*

---

Rose S. Lee